DOCUMENT RESUME

ED 418 992		TM 028 276

AUTHOR	Henard, David H.
TITLE	Using Spreadsheets To Implement the One-Parameter Item Response Theory (IRT) Model.
PUB DATE	1998-04-11
NOTE	38p.; Paper presented at the Annual Meeting of the Southwestern Psychological Association (New Orleans, LA, April 1998). Small print in tables/figures may not reproduce clearly.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Computer Oriented Programs; Computer Software; Heuristics; *Item Response Theory; *Mathematical Models; *Spreadsheets; *Test Items
IDENTIFIERS	*One Parameter Model; *Rasch Model

ABSTRACT

Item response theory models arose from the inherent limitations of classical test theory methods of test analysis. A brief description of those limitations and the corresponding enhancements provided by item response models is provided. Further, an examination of the popular Rasch one-parameter latent trait model is undertaken. Specific explanation of the step-by-step calculations in the one-parameter model is accomplished using a commonly available spreadsheet. This paper is designed to be used as a teaching heuristic to assist students in understanding both the mechanics and the rationale behind the item response theory model measurement. (Contains nine tables, four figures, and nine references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 418 992

Using Spreadsheets to Implement the One-Parameter Item Response Theory (IRT) Model

David H. Henard
Texas A&M University
Mays Graduate School of Business
College Station, TX 77843-4112
409.845.5205 (phone)
409.862.2811 (fax)
henard@tamu.edu

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
David Henard

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM028276

Paper presented at the annual meeting of the Southwestern Psychological Association, New Orleans, April 11, 1998.

BEST COPY AVAILABLE

ABSTRACT

Item response theory models arose from the inherent limitations of classical test theory methods of test analysis. A brief description of those limitations and the corresponding enhancements provided by item response models is provided. Further, an examination of the popular Rasch one-parameter latent trait model is undertaken. Specific explanation of the step-by-step calculations in the one-parameter model is accomplished using a commonly available spreadsheet. This paper is designed to be used as a teaching heuristic to assist students in understanding both the mechanics and the rationale behind the item response theory (IRT) model measurement.

The author would like to express his thanks to Bruce Thompson for his comments on an earlier draft of this paper.

When they were first introduced, item response theory (IRT)/latent trait measurement models were heralded as "one of the most important methodological advances in psychological measurement in the past half century" (McKinley and Mills 1989, p. 71). However, the pluses and minuses of these models have been hotly debated (cf. Lawson 1991) despite their widespread use in various applications such as test equating, item selection and adaptive testing.

This paper will begin with a brief examination of classical test theory and some of its inherent weaknesses. An encompassing examination of classical test theory is beyond the scope of the paper. Readers desiring greater discourse on the subject are directed to Crocker and Algina (1986) and Nunnelly and Bernstein (1994). The focus then shifts to item response theory with discussion centering on the theoretical framework of the Rasch one-parameter IRT model. The concepts underlying and the basic tenets of item response theory are explored. Finally, step-by-step calculations involved in the Rasch model will be explained using a commonly available spreadsheet. Such spreadsheets can be valuable heuristic devises to assist students in truly understanding what is occurring in IRT measurement.

Classical Test Theory

Classical test theory (CTT) and its related methods has a number of limitations. For example, comparison across examinees is limited to situations where the subjects of interest are administered the same (parallel) test items. Also, a false presumption of CTT is that the variance of errors of measurement is the same for all examinees. Reality

ERIC

dictates that some people perform tasks more consistently than others and that consistency varies with ability (Hambleton and Swaminathan 1985).

Two major CTT limitations of note are:

1. *Examinee characteristics cannot be separated from test characteristics,* and
2. *CTT is test-oriented rather than item-oriented.*

The first limitation above can be summarized as a situation of circular dependency. The examinee statistic (i.e., observed score) is item sample-dependent while the item statistics (i.e., item difficulty, item discrimination) are examinee sample-dependent. Stated simply, when the test is 'difficult', examinees will appear to have lower ability and when the test is 'easy', they will appear to have higher ability. Likewise, the 'difficulty' of a test item is determined by the proportion of examinees who answer it correctly and is thus dependent on the abilities of the examinees being measured (Hambleton, Swaminathan and Rogers 1991). This circular dependency poses some theoretical difficulties in CTT's application in measurement situations such as test equating and computerized adaptive testing.

The second major limitation listed is a question of orientation. The CTT model fails to allow us to predict how an examinee, given a stated ability level, is likely to respond to a particular item (Hambleton et al. 1991). Predicting how an individual examinee or a group of examinees will perform on a specific item is quite relevant to a number of testing applications. Consider the difficulties facing a test designer who wishes to predict test scores across multiple groups, or to design an equitable test for a particular group, or possibly to compare examinees who take either different tests or the same test at differing times. Such inherent limitations of CTT led psychometricians to develop models that overcame not only these limitations but also led to improved bias detection, enhanced reliability assessment and increased precision in ability measurement. Item response theory (Hambleton and Swaminathan 1985; Hambleton et al. 1991; Lord 1980) provides us with a framework to accomplish these desired features.

Item Response Theory Concept

Item response theory (IRT) arose out of a psychometric need to overcome the limitations of classical test theory and to provide test designers with improved and more accurate testing tools. Again, a thorough discussion of IRT is beyond the bounds of the present study. Interested readers are directed to Hambleton et al. (1991) and Wright and Stone (1979). IRT primarily rests upon two basic postulates (Hambleton et al. 1991; Hambleton and Swaminathan 1985):

1. The performance of an examinee on a test item can be explained (or predicted) by a set of factors called *traits*, *latent traits* or *abilities*; and
2. The relationship between examinees' item performance and the trait(s) underlying item performance can be described by a monotonically increasing function called the *item characteristic curve* (ICC).

Several IRT models exist, including the three-parameter, two parameter and one-parameter models. The one-parameter model, often referred to as the Rasch model, is the most commonly used example and will be the focus of this paper. The models differ principally in the mathematical form of the ICC and/or the number of parameters specified in the model.

ERIC

When an IRT model fits the test data of interest, many of the limitations of CTT are resolved. For example, examinee latent trait estimates are theoretically no longer test-dependent and item indices are no longer group-dependent. Ability estimates derived from different groupings of items will be the same, barring measurement error, and item parameter estimates derived from different groups of examinees will also be the same, barring sampling error (Hambleton et al. 1991).

Assumptions of Item Response Models

Unidimensionality and *local independence* are two assumptions that are fundamental to IRT. The unidimensionality assumption requires that only one ability or latent trait is measured by the various items that make up the test. Intuitively, this assumption cannot be strictly satisfied due to the reality that multiple factors will normally impact the test taking performance by an examinee. Exogenous factors such as generic cognitive ability, test anxiety and motivation level are likely to impact test performance as well. In order for a set of test data to satisfy the assumption of unidimensionality, a 'dominant' factor influencing performance must be present (Hambleton et al. 1991). This dominant factor is referred to as the ability or latent trait measured by the test.

The assumption of local independence requires that an examinee's responses to the various items in a test are statistically independent of each other (Hambleton and Swaminathan 1985). This implies that an examinee's response to any one item will not affect their response to any other item in the test. Simply put, the trait specified in the model is the only factor influencing the respondent's answer to the test items (Hambleton et al. 1991) and one item does not hold clues for subsequent items. It is important to note that the assumption of local independence does not imply that the test items are uncorrelated across the total group of examinees (Lord and Novak 1968). Whenever there is variation among the examinees on the measured ability, positive correlations between pairs of items will result. However, item scores are uncorrelated at a fixed ability level (Hambleton and Swaminathan 1985).

There are three primary advantages to using item response models (Hambleton and Swaminathan 1985):

1. Assuming the existence of a large pool of items each measuring the same latent trait, the estimate of an examinee's ability is independent of a particular sample of test items that are administered to the examinee;
2. Assuming the existence of a large population of examinees, the descriptors of a test item (e.g., item difficulty, item discrimination) are independent of the particular sample of examinees drawn for the purpose of item calibration; and
3. A statistic indicating the precision with which each examinee's ability is estimated is provided.

Thus, the primary argument for employing IRT methods is that the resulting analyses are both person-free and sample-free measurements (McKinley and Mills 1989). It should be noted that not all researchers agree that IRT offers us such rich benefits. Lawson (1991) subjected three test data sets to both classical and Rasch procedures and found "remarkable similarities" between the results. Findings for both examinee abilities and item difficulties yielded "almost identical information." Given the mathematical intricacies of IRT that are not required of classical methods, Lawson questioned the necessity of the Rasch procedure. That is, once misfitting items and people are removed from the analysis, IRT and CTT models seem to yield highly correlated person ability and item difficulty estimates. The Rasch model continues, however, to be utilized by psychometricians. The recent rise in adaptive testing bears testament to the continued use of IRT.

THE RASCH MODEL CALCULATIONS

Having noted the basic deficiencies of classical test theory and the improvements that the more theoretically-based item response theory provides us, attention is now focused on the step-by-step calculations in the one-parameter IRT measurement. The Rasch calculations can appear daunting to many students. While extremely powerful in its applications, the fundamentals of IRT are actually quite straightforward and should not be viewed as a *black box* process. It is hoped that the following discussion will facilitate the conceptual grasp of the subject.

In the following data example, presume that 35 people were tested on an 18 item exam. Since the object of the item response model is to predict performance based on item calibrations that are independent of the persons generating the data (i.e., person free) and examinee ability estimates that are independent of the items used in the measurement (i.e., item free), all items that are answered either correctly or incorrectly by everyone will be removed from further analysis. Likewise, any person who answered either 0% or 100% of the items correctly will also be removed since neither can be calibrated against the group and thus provide us with no usable information. That is, such items and people provide no information to facilitate the estimation process (e.g., the person with all of the items correct may be exactly smart enough to do that, or may have any of the infinite ability levels above the exact ability that is just sufficient to yield this perfect score. The resulting data set after this initial cut of the information can be seen in Table 1.

Insert Table 1 about here

Table 1 is laid out so that examinees are sorted in increasing order of number of items answered correctly while items are sorted in increasing order of number of examinees that correctly answered the item. Since responses were dichotomously scored as either right or wrong, a '0' in the table denotes an incorrect answer while a '1' denotes a correct response. Looking at Table 1, examinee 25 answered the fewest number of questions correctly (2) while examinees 24, 34 and 7 answered the greatest number of items correctly (11). Remember that any examinee who scored perfectly 100 or zero has been removed. Any 'perfect' items have also been removed. In this data set, item numbers 1,2,3 and 18 were removed while examinee 35 was removed. This editing of the data continues in this manner until no 'perfect' items or persons remain.

Given this initial editing of the data, the next step in the process is to calibrate both the item difficulties and the person abilities. In order for us to make valid assessments and predictions arising from the Rasch model, both of these statistics (difficulties and abilities) must be linear and in the same metric. In IRT, this is accomplished by converting the values into *logits*. Logits for item difficulties are calculated as the natural log of the proportion of items incorrect divided by the proportion correct.[1] Conversely, the logit calculation for person ability is the natural log of the proportion of items that an examinee correctly answered divided by the proportion

[1] $ln\,[(1-p_i)\,/\,p_i]$

answered incorrectly.[2] These conversions from proportions to logits can be seen in greater detail in Tables 2 and 3, respectively.

Insert Tables 2 & 3 about here

Once the logit values for both the persons and items are calculated, we have overcome another weakness associated with CTT. Namely, while item difficulty and person ability levels realistically range from negative infinity to positive infinity, the proportion correct/incorrect are bound by the values of zero and one. Conversion to logits transforms the values into a +/- ∞ scale. One further step involves calculating the mean and standard deviation of the data and converting the logits to a standard (z) scale arbitrarily assigning a center point value of zero. While the scale theoretically runs from -∞ to + ∞, values realistically tend to vary between +/- 3 logits. Table 4 highlights the relationship between proportion (of correct responses) and personal ability logits for this data set. Additionally, Figure 1 graphically portrays the relationship between proportions and logits.

Insert Table 4 and Figure 1 about here

Two final calibration steps remain. The initial measurement of item difficulty must be corrected for the difficulty dispersion of the items. Additionally, the initial measurement of person ability needs to be corrected for the ability dispersion of persons. Calculations are modeled in Tables 5 and 6 and result in item calculations that are corrected for sample spread and person calculations that are corrected for test width. This

[2] $ln\,[p_i / (1-p_i)]$

step is known as the calculation of *expansion factors* and is crucial given the premise of one-parameter IRT that the achievement of any person on a given item is solely dependent upon that person's ability and the difficulty of the specific item.

Insert Tables 5 & 6 about here

The final step in the modeling process is to fit the model, obtained via the preceding steps, to the data and evaluate the goodness of fit. One can not merely assume that the preceding steps are sufficient in developing an effective model. If we re-examine Table 1, a pattern in responses should emerge. Since items are ordered by increasing level of difficulty and examinees are progressively ranked according to correct responses, we would intuitively expect to see more incorrect responses to the top and right of the table and vice versa. In other words, we would expect a person who is higher on a latent trait (Θ, theta) to have a greater chance of answering a difficult question than a person who is lower on that latent trait. Table 7 accentuates the different 'expectations'.

While those persons or items that do not fit well with the model are statistically identified by the software program used to calculate the Rasch model, some of the potential misfits are circled here in order to visually highlight points where the model does not perfectly fit the data. It should be noted that both items and persons can be identified as aberrant. For example, person 13 answered item 12 correctly when the expectation (given other responses) would be that the item would be answered incorrectly. Also, items 6 and 8 were answered incorrectly by person 12 when the expectation would be a correct response.

Insert Table 7 about here

Once an IRT computer program identifies misfits between the model and the data, the source of the variance (item- or person-based) is explored. In examining Table 7, persons 13 and 29 appear to post responses that are aberrant to expectations. Likewise, items 6, 7, 8 and 12 do not fit with expectations. Each of these variants in the table are circled for easier identification. As mentioned previously, the source of the inconsistencies can originate at either the item or person level. Table 8 simulates how the software program would investigate the irregularities caused by persons 29 and 13, for example.

Each item has a calculated difficulty level (d) and each person has a calculated ability level (theta). The first step in the analysis of fit is to determine the difference between the ability level and the difficulty level for each person and each item. Table 8 highlights the values involved for persons 29 and 13. When the difference in the two values is a positive number, it is an indication that that particular item should be 'easy' for that particular examinee and should be answered correctly. The higher the number, the greater the likelihood of a correct response. Conversely, the more negative the difference, the greater the likelihood that the item difficulty exceeds the person's ability.

Insert Table 8 about here

Looking at Table 8, it appears that person 29 missed item 7, when they should have theoretically answered it correctly, while correctly answering item 14, when the probability was that it would be missed by a person with a theta equal to zero. Person 13 missed both items 6 and 7 while getting item 12 correct--all opposite of expectations. Remember, however, that the source of variance can result from item irregularities as well. Table 9 facilitates our understanding of how item response patterns are examined for misfitting results. The process is similar to the aforementioned one. This table illustrates the examination of items 6 and 7 for all persons. Again, the process occurs for all items and all persons.

Insert Table 9 about here

Both items 6 and 7 have fairly high negative logit values for item difficulty levels indicating that they should be answered correctly by most examinees. Indeed, an examination of the results in Table 9 shows that only three of the 34 examinees missed item 7 while only four missed item 6. It appears that it is not items 6 and 7 that are causing the irregularity between the model and the data but rather examinees 13 and 29. This is exactly what is occurring. The removal of these two persons from the data set eliminates most of the irregularity associated with the two items.

Upon removal of persons 13 and 29, the process iterates and a new evaluation of fit is calculated for the remaining distributions. Again, all combinations of persons and items are examined. At the point at which no further removal of either items or persons enhances the goodness of fit, the model is said to "fit" the data and the result is items that are theoretically both unidimensional and independent. By eliminating both items and individuals that deviate from expectations, we can develop a test bank of items that should optimally fit the individual person ability levels for most test takers.

Figure 2 illustrates one-parameter item characteristic curves (ICC) for four hypothetical items. Latent ability (Θ) is represented, in logits, along the x-axis. The probability of a correct response is located on the y-axis. Since in the one-parameter IRT model no traits other than ability (e.g., guessing) are assumed to impact responses, the curves are asymptotic to the zero and one points of the probability distribution. The difficulty level of each item is defined as the logit point at which the probability of answering the item correctly is 50% ($p = 0.50$). Therefore, those items with curves that are toward the right side of x-axis are more difficult than those to the left. For example, the item difficulty for item 3 is -1.0 while the item difficulty for item 2 is approximately +2.0. Therefore, persons with an ability (Θ) equal to zero would probably answer item 3 correctly and miss items 1 and 2. There is a 50% chance of the person answering item 4 correctly.

Insert Figure 2 about here

Figures 3 and 4 are simply added as a point of comparison and for further edification. The two-parameter model assumes two parameters are affecting examinee responses: ability and item discrimination. Curve endpoints are still asymptotic as answers can only be correct or incorrect. With the two-parameter curve, the slope of the curve indicates how well the item differentiates between persons with varying latent abilities. For instance, item 2 in Figure 3 has a much flatter slope than that of item 4. Therefore, item 4 is a better discriminating item.

Figure 4, the three-parameter model, adds a third variable to the equation--the effect of guessing. Here, the curve endpoint may begin at a value other than zero as the impact of correctly guessing an item is taken into account. The evaluation guidelines that applied to the other two ICCs apply here as well; however, the location of the initial endpoint gives the researcher an indication as to how effective item distracters may be. For example, items 3 and 6 appear to be potentially guessed correctly whereas items 2 and 4 do not.

Insert Figures 3 and 4 about here

SUMMARY

Item response theory models arose from the inherent limitations of classical test theory methods of test analysis. Chief among the limitations is that examinee characteristics can not be separated from test characteristics. Item response theory overcomes these limitations and rests on two major assumptions: (a) the performance of an examinee can be explained by a set of factors known as traits, and (b) the relationship between an individual's item performance can be described by a monotonically increasing function termed an item characteristic curve.

Item response theory allows the researcher to develop test questions that are theoretically both person-free and item-free. IRT stresses maximizing the test information function over the range of abilities that are of interest instead of maximizing reliability, as does classical psychometrics. While the usefulness of IRT continues to be debated, IRT appears to hold many benefits. Among these are a more accurate ability to detect item or test bias, the ability to administer customized, individualized, computer-adaptive tests and the ability to construct more effective tests, in general. It is hoped that

ERIC

this paper has facilitated a better understanding of both the mechanics and the rationale behind item response theory (IRT) measurement.

ERIC

REFERENCES

Crocker, L. and J. Algina (1986), *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.

Hambleton, Ronald K. and H. Swaminathan (1985), *Item Response Theory: Principles and Applications*. Boston: Kluwer.

----, ----, and H. Jane Rogers (1991), *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications.

Lawson, Stephen (1991), "One Parameter Latent Trait Measurement: Do the Results Justify the Effort?," in *Advances in Education Research: Substantive Findings, Methodological Developments*, Bruce Thompson, ed. Greenwich: JAI Press.

Lord, F. M. (1980), *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum.

---- and M. R. Novak (1968), *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

McKinley, R. and C. Mills (1989), "Item Response Theory: Advances in Achievement and Attitude Measurement," in *Advances in Social Science Methodology*, Bruce Thompson, ed. Greenwich: JAI Press.

Nunnally, Jum C. and Ira H. Bernstein (1994), *Psychometric Theory*. New York: McGraw-Hill, Inc.

Wright, Benjamin D. and Mark H. Stone (1979), *Best Test Design*. Chicago: MESA Press.

BEST COPY AVAILABLE

Table 1

Edited Responses of 34 Examinees

	Item #															
Person	4	5	7	6	9	8	10	11	13	12	14	15	16	17	Person Score	P of 14
25	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2	0.14
4	1	0	1	0	1	0	0	0	0	0	0	0	0	0	3	0.21
33	1	0	1	0	0	0	1	0	0	0	0	0	0	0	3	0.21
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	4	0.29
27	1	1	1	1	0	0	0	0	0	0	0	0	0	0	4	0.29
11	0	1	1	1	1	1	0	0	0	0	0	0	0	0	5	0.36
12	1	1	1	0	1	0	1	0	0	0	0	0	0	0	5	0.36
17	1	0	1	1	1	1	1	0	0	0	0	0	0	0	6	0.43
19	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0.43
30	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0.43
2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50
3	1	1	1	1	1	1	0	0	0	1	0	0	0	0	7	0.50
5	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50
6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50
8	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50
9	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50
13	1	1	0	0	1	1	1	1	0	1	0	0	0	0	7	0.50
16	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0.50
26	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50
28	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0.50
29	1	1	0	1	1	0	1	1	0	0	1	0	0	0	7	0.50
31	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50
10	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0.57
18	1	1	1	1	1	1	1	0	1	0	0	0	0	0	8	0.57
14	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0.57
32	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0.57
20	1	1	1	1	1	1	1	0	1	0	0	0	0	0	8	0.57
21	1	1	1	1	1	1	1	1	1	0	0	0	0	0	9	0.64
22	1	1	1	1	1	1	1	1	0	1	0	0	0	0	9	0.64
23	1	1	1	1	1	1	1	0	1	0	1	0	0	0	9	0.64
34	1	1	1	1	1	1	1	0	0	1	1	0	0	0	9	0.64
15	1	1	1	1	1	1	1	1	1	1	0	0	0	0	10	0.71
7	1	1	1	1	1	1	1	1	1	1	0	1	0	0	11	0.79
24	1	1	1	1	1	1	1	1	1	0	0	0	1	1	11	0.79
34	1	1	1	1	1	1	1	1	1	0	0	0	1	1	11	0.79
Item Score	32	31	31	30	30	27	24	12	7	6	3	1	1	1		
P of 34	0.94	0.91	0.91	0.88	0.88	0.79	0.71	0.35	0.21	0.18	0.09	0.03	0.03	0.03		

19

20

Table 2

Distribution of Item Scores

Person	Item	Item Score	Freq.	P_i	$1-P_i$	$1-P_i / P_i$	Logit Ratio (x_i)	Freq. * Logit	x_i^2	Freq. * x_i^2	Logit-mean (d)	Freq. * x_i^2
1	4	32	1	0.941	0.059	0.063	-2.77	-2.77	7.69	7.69	-2.96	0.035
2	5,7	31	2	0.912	0.088	0.097	-2.34	-4.67	5.45	10.91	-2.52	0.070
3	6,9	30	2	0.882	0.118	0.133	-2.01	-4.03	4.06	8.12	-2.20	0.070
4	8	27	1	0.794	0.206	0.259	-1.35	-1.35	1.82	1.82	-1.54	0.035
5	10	24	1	0.706	0.294	0.417	-0.88	-0.88	0.77	0.77	-1.06	0.035
6	11	12	1	0.353	0.647	1.833	0.61	0.61	0.37	0.37	0.42	0.035
7	13	7	1	0.206	0.794	3.857	1.35	1.35	1.82	1.82	1.16	0.035
8	12	6	1	0.176	0.824	4.667	1.54	1.54	2.37	2.37	1.35	0.035
9	14	3	1	0.088	0.912	10.333	2.34	2.34	5.45	5.45	2.15	0.035
10	15-17	1	3	0.029	0.971	33.000	3.50	10.49	12.23	36.68	3.31	0.105

N = 34

BEST COPY AVAILABLE

Table 3

Distribution of Person Scores

Possible Right	Person Freq.	P Correct (P_i)	$1-P_i$	$P_i / 1-P_i$	Logit Correct	Freq. * Logit	Logit^2	Freq. * Logit^2
1	0	0.07	0.93	0.077	-2.56	0.00	6.579	0.000
2	1	0.14	0.86	0.167	-1.79	-1.79	3.210	3.210
3	2	0.21	0.79	0.273	-1.30	-2.60	1.688	3.376
4	2	0.29	0.71	0.400	-0.92	-1.83	0.840	1.679
5	2	0.36	0.64	0.556	-0.59	-1.18	0.345	0.691
6	3	0.43	0.57	0.750	-0.29	-0.86	0.083	0.248
7	12	0.50	0.50	1.000	0.00	0.00	0.000	0.000
8	5	0.57	0.43	1.333	0.29	1.44	0.083	0.414
9	4	0.64	0.36	1.800	0.59	2.35	0.345	1.382
10	1	0.71	0.29	2.500	0.92	0.92	0.840	0.840
11	2	0.79	0.21	3.667	1.30	2.60	1.688	3.376
12	0	0.86	0.14	6.000	1.79	0.00	3.210	0.000
13	0	0.93	0.07	13.000	2.56	0.00	6.579	0.000

L = 14

BEST COPY AVAILABLE

Table 4
Logit Conversion Chart

Proportion	Logit	Proportion	Logit	Proportion	Logit	Proportion	Logit
0.01	-4.60	0.26	-1.05	0.51	0.04	0.76	1.15
0.02	-3.89	0.27	-0.99	0.52	0.08	0.77	1.21
0.03	-3.48	0.28	-0.94	0.53	0.12	0.78	1.27
0.04	-3.18	0.29	-0.90	0.54	0.16	0.79	1.32
0.05	-2.94	0.30	-0.85	0.55	0.20	0.80	1.39
0.06	-2.75	0.31	-0.80	0.56	0.24	0.81	1.45
0.07	-2.59	0.32	-0.75	0.57	0.28	0.82	1.52
0.08	-2.44	0.33	-0.71	0.58	0.32	0.83	1.59
0.09	-2.31	0.34	-0.66	0.59	0.36	0.84	1.66
0.10	-2.20	0.35	-0.62	0.60	0.41	0.85	1.73
0.11	-2.09	0.36	-0.58	0.61	0.45	0.86	1.82
0.12	-1.99	0.37	-0.53	0.62	0.49	0.87	1.90
0.13	-1.90	0.38	-0.49	0.63	0.53	0.88	1.99
0.14	-1.82	0.39	-0.45	0.64	0.58	0.89	2.09
0.15	-1.73	0.40	-0.41	0.65	0.62	0.90	2.20
0.16	-1.66	0.41	-0.36	0.66	0.66	0.91	2.31
0.17	-1.59	0.42	-0.32	0.67	0.71	0.92	2.44
0.18	-1.52	0.43	-0.28	0.68	0.75	0.93	2.59
0.19	-1.45	0.44	-0.24	0.69	0.80	0.94	2.75
0.20	-1.39	0.45	-0.20	0.70	0.85	0.95	2.94
0.21	-1.32	0.46	-0.16	0.71	0.90	0.96	3.18
0.22	-1.27	0.47	-0.12	0.72	0.94	0.97	3.48
0.23	-1.21	0.48	-0.08	0.73	0.99	0.98	3.89
0.24	-1.15	0.49	-0.04	0.74	1.05	0.99	4.60
0.25	-1.10	0.50	0.00	0.75	1.10		

BEST COPY AVAILABLE

Table 5

Final Estimates of Item Difficulty

Possible Right	Item Name	Initial Item Calibration	Spread Expansion Factor	Corrected Item Calibration	Item Score	Standard Error
1	4	-2.94	1.31	-3.85	32	0.95
2	5,7	-2.50	1.31	-3.28	31	0.79
3	6,9	-2.18	1.31	-2.86	30	0.70
4	8	-1.51	1.31	-1.98	27	0.56
5	10	-1.09	1.31	-1.43	24	0.49
6	11	0.43	1.31	0.56	12	0.47
7	13	1.13	1.31	1.48	7	0.56
8	12	1.33	1.31	1.74	6	0.59
9	14	2.12	1.31	2.78	3	0.79
10	15,16,17	3.29	1.31	4.31	1	1.33

N = 34

BEST COPY AVAILABLE

ERIC

Table 6
Final Estimates of Person Ability

Possible Right	Initial Measure	Test Width Expansion factor	Corrected Measure	Standard Error
1	-2.59	2.09	-5.41	2.17
2	-1.82	2.09	-3.80	1.60
3	-1.32	2.09	-2.76	1.36
4	-0.90	2.09	-1.88	1.24
5	-0.58	2.09	-1.21	1.17
6	-0.28	2.09	-0.59	1.13
7	0.00	2.09	0.00	1.12
8	0.28	2.09	0.59	1.13
9	0.58	2.09	1.21	1.17
10	0.90	2.09	1.88	1.24
11	1.32	2.09	2.76	1.36
12	1.82	2.09	3.80	1.60
13	2.59	2.09	5.41	2.17

L = 14

BEST COPY AVAILABLE

Table 7
Response Patterns of Persons to Items

		Easier Items							Harder Items									
	Person	4	5	7	6	9	8	10	11	13	12	14	15	16	17	Person Score	P of 14	Ability Logits
	25	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2	0.14	-3.8
	4	1	0	1	0	1	0	0	0	0	0	0	0	0	0	3	0.21	-2.8
	33	1	0	1	0	0	0	1	0	0	0	0	0	0	0	3	0.21	-2.8
	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	4	0.29	-1.9
Low Persons	27	1	1	1	1	0	0	0	0	0	0	0	0	0	0	4	0.29	-1.9
	11	0	1	1	1	1	1	0	0	0	0	0	0	0	0	5	0.36	-1.2
	12	1	1	1	0	1	0	1	0	0	0	0	0	0	0	5	0.36	-1.2
	17	1	0	1	1	1	1	1	0	0	0	0	0	0	0	6	0.43	-0.6
	19	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0.43	-0.6
	30	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	0.43	-0.6
	2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50	0
	3	1	1	1	1	1	1	0	0	0	1	0	0	0	0	7	0.50	0
	5	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50	0
	6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50	0
	8	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50	0
	9	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50	0
	13	1	1	0	0	1	1	1	1	0	1	0	0	0	0	7	0.50	0
	16	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0.50	0
	26	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50	0
	28	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0.50	0
	29	1	1	0	1	1	0	1	1	0	0	1	0	0	0	7	0.50	0
	31	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0.50	0
	10	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0.57	0.6
	18	1	1	1	1	1	1	1	0	1	0	0	0	0	0	8	0.57	0.6
	14	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0.57	0.6
	32	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0.57	0.6
High Persons	20	1	1	1	1	1	1	1	0	1	0	0	0	0	0	8	0.57	0.6
	21	1	1	1	1	1	1	1	1	1	0	0	0	0	0	9	0.64	1.2
	22	1	1	1	1	1	1	1	1	0	1	0	0	0	0	9	0.64	1.2
	23	1	1	1	1	1	1	1	0	1	0	1	0	0	0	9	0.64	1.2
	34	1	1	1	1	1	1	1	0	0	1	1	0	0	0	9	0.64	1.2
	15	1	1	1	1	1	1	1	1	1	1	0	0	0	0	10	0.71	1.9
	7	1	1	1	1	1	1	1	1	1	1	0	0	0	0	11	0.79	2.8
	24	1	1	1	1	1	1	1	1	1	0	0	1	0	0	11	0.79	2.8
	34	1	1	1	1	1	1	1	1	1	0	0	0	1	1	11	0.79	2.80
	Item Score	32	31	31	30	30	27	24	12	7	6	3	1	1	1			
	P of 34	0.94	0.91	0.91	0.88	0.88	0.79	0.71	0.35	0.21	0.18	0.09	0.03	0.03	0.03			
	Item Diff. Logits	-3.90	-3.30	-3.30	-2.90	-2.90	-2.00	-1.40	0.60	1.50	1.70	2.80	4.30	4.30	4.30			

31

BEST COPY AVAILABLE

Table 8
Fit Analysis for Persons 29 and 13

			Item 4	5	7	6	9	8	10	1	13	12	14	15	16	17	Number Right	Sum of Squares
		Difficulty (d)	-3.9	-3.3	-3.3	-2.9	-2.9	-2.0	-1.4	0.6	1.5	1.7	2.8	4.3	4.3	4.3		
Person 29 (theta = θ)		Response:	1	1	0	1	1	0	1	1	0	0	1	0	0	0	7	
	x = 0	(theta - d)			3.3			2.0			-1.5	-1.7		-4.3	-4.3	-4.3		
	x = 1	(d - theta)	-3.9	-3.3		-2.9	-2.9		-1.4	0.6			2.8					
	z^2		0.0	0.0	27.0	0.0	0.0	7.0	0.0	2.0	0.0	0.0	17.0	0.0	0.0	0.0		53.0
Person 13 (theta = θ)		Response:	1	1	0	0	1	1	1	1	0	1	0	0	0	0	7	
	x = 0	(theta - d)			3.3	2.9					-1.5		-2.8	-4.3	-4.3	-4.3		
	x = 1	(d - theta)	-3.9	-3.3			-2.9	-2	-1.4	0.6		1.7						
	z^2		0.0	0.0	27.0	18.0	0.0	0.0	0.0	2.0	0.0	6.0	0.0	0.0	0.0	0.0		53.0

BEST COPY AVAILABLE

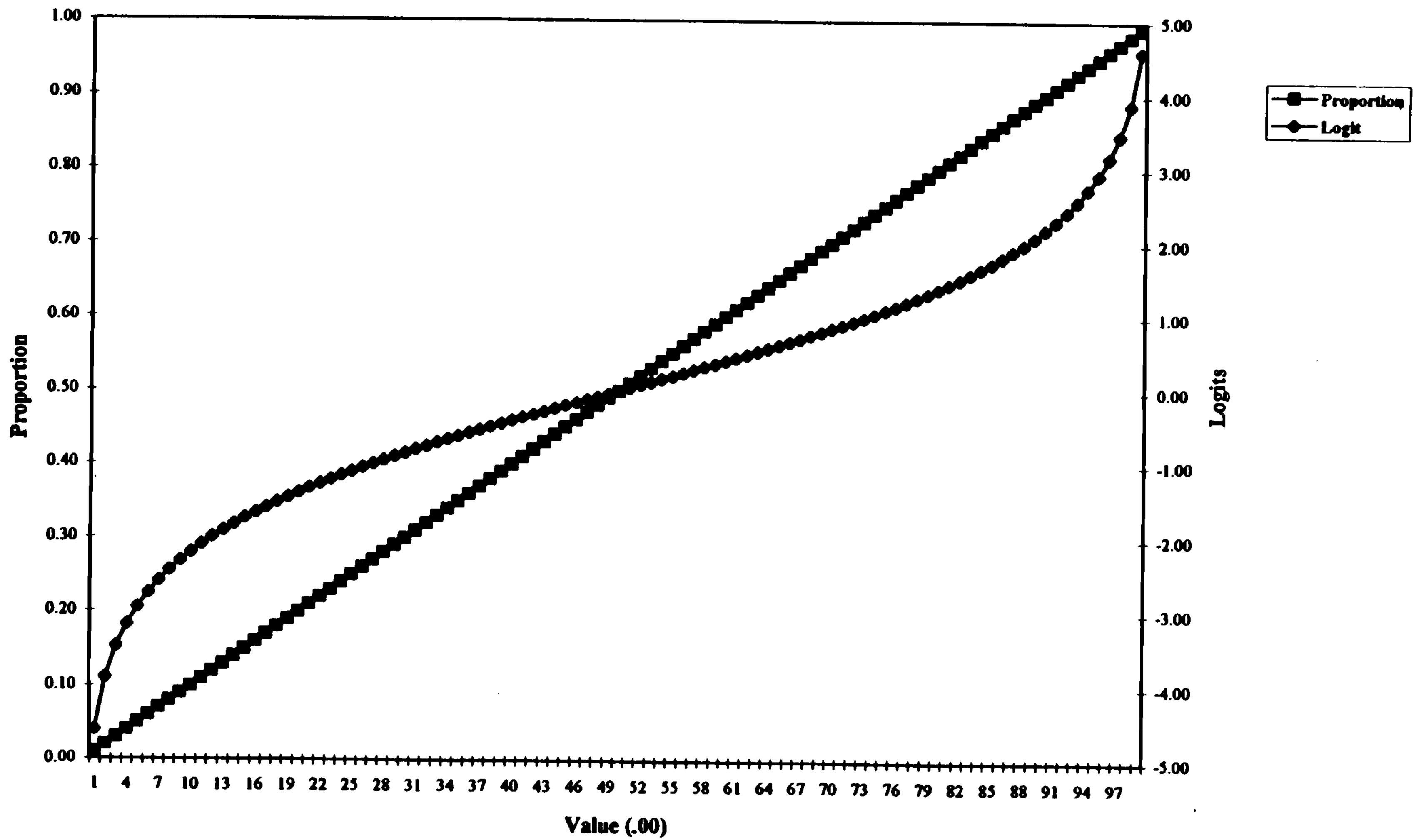
Table 9
Fit Analysis for Items 6 and 7

Person	Ability (theta)	Item 7 (d = -3.3) Response	x = 0 (theta-d)	x = 1 (d-theta)	z^2	Item 6 (d = -2.9) Response	x = 0 (theta-d)	x = 1 (d-theta)	z^2
25	-3.8	0	-0.5		1	1		0.9	3
4	-2.8	1		-0.5	1	0	0.1		1
33	-2.8	1		-0.5	1	0	0.1		1
1	-1.9	1		-1.4	0	1		-1	0
27	-1.9	1		-1.4	0	1		-1	0
11	-1.2	1		-2.1	0	1		-1.7	0
12	-1.2	1		-2.1	0	0	1.7		6
17	-0.6	1		-2.7	0	1		-2.3	0
19	-0.6	1		-2.7	0	1		-2.3	0
30	-0.6	1		-2.7	0	1		-2.3	0
2	0	1		-3.3	0	1		-2.9	0
3	0	1		-3.3	0	1		-2.9	0
5	0	1		-3.3		1		-2.9	0
6	0	1		-3.3	0	1		-2.9	0
8	0	1		-3.3	0	1		-2.9	0
9	0	1		-3.3	0	1		-2.9	0
13	0	0	3.3		27	0	2.9		18
16	0	1		-3.3	0	1		-2.9	0
26	0	1		-3.3	0	1		-2.9	0
28	0	1		-3.3	0	1		-2.9	0
29	0	0	3.3		27	1		-2.9	0
31	0	1		-3.3	0	1		-2.9	0
10	0.6	1		-3.9	0	1		-3.5	0
18	0.6	1		-3.9	0	1		-3.5	0
14	0.6	1		-3.9	0	1		-3.5	0
32	0.6	1		-3.9	0	1		-3.5	0
20	0.6	1		-3.9	0	1		-3.5	0
21	1.2	1		-4.5	0	1		-4.1	0
22	1.2	1		-4.5	0	1		-4.1	0
23	1.2	1		-4.5	0	1		-4.1	0
34	1.2	1		-4.5	0	1		-4.1	0
15	1.9	1		-5.2	0	1		-4.8	0
7	2.8	1		-6.1	0	1		-5.7	0
24	2.8	1		-6.1	0	1		-5.7	0
Sum of Squares					57				29

BEST COPY AVAILABLE

Figure 1

Scatterplot of Proportions to Logits



BEST COPY AVAILABLE

Figure 2
One-Parameter Model

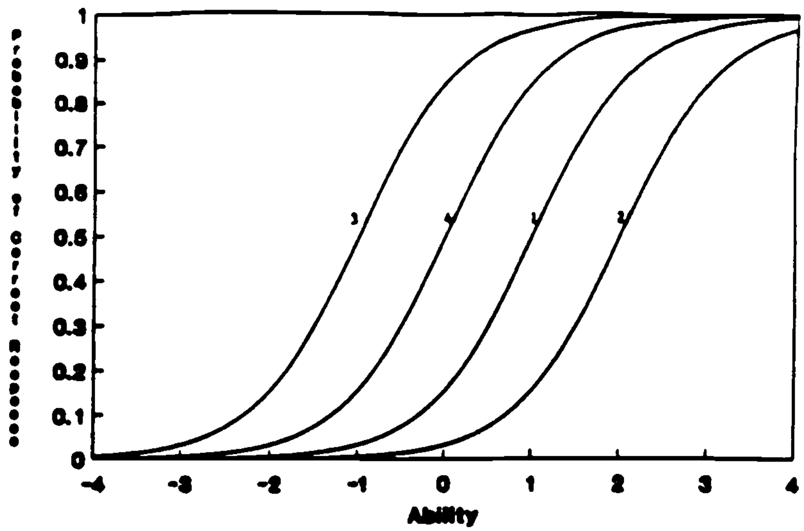


Figure 3
Two-Parameter Model

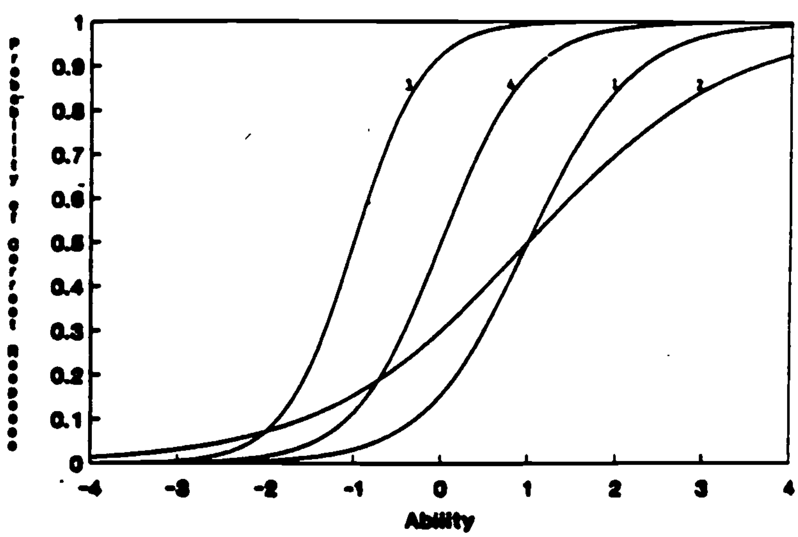
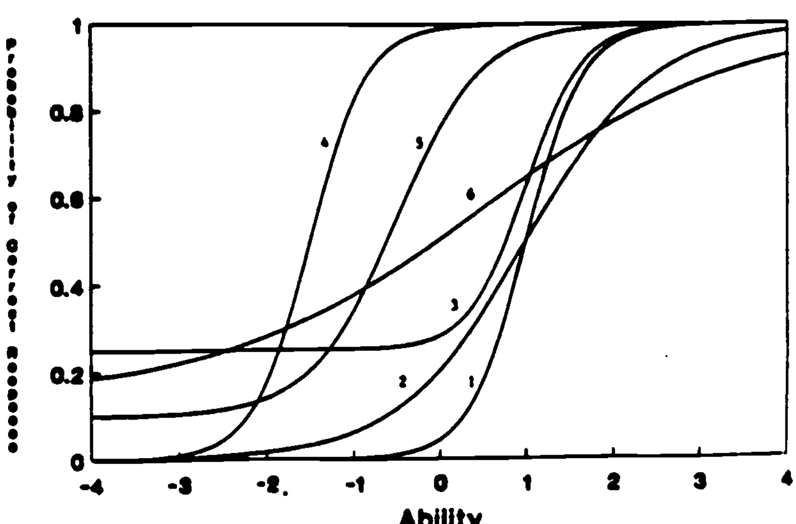


Figure 4
Three-Parameter Model

BEST COPY AVAILABLE



TM028276

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: USING SPREADSHEETS TO IMPLEMENT THE ONE-PARAMETER ITEM RESPONSE THEORY (IRT) MODEL	
Author(s): DAVID H. HENARD	
Corporate Source:	Publication Date: 4/11/98

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

X Sample sticker to be affixed to document | Sample sticker to be affixed to document ☐

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
DAVID H. HENARD
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here
Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X David H. Henard	Position: RES ASSOCIATE
Printed Name: DAVID H. HENARD	Organization: TEXAS A&M UNIVERSITY
Address: TAMU GRAD SCH OF BUSINESS COLLEGE STATION, TX 77843-4112	Telephone Number: (409)845-1335
	Date: 3/23/98

ERIC